

BUNDESREPUBLIK DEUTSCHLAND

EP 00/4803



**PRIORITY
DOCUMENT**
SUBMITTED OR TRANSMITTED IN
COMPLIANCE WITH RULE 17.1(a) OR (b)

**Prioritätsbescheinigung über die Einreichung
einer Patentanmeldung**

REC'D 23 OCT 2000

WIPO

PCT

Aktenzeichen:

199 25 046.4

Anmeldetag:

01. Juni 1999

Anmelder/Inhaber:

Alcatel, Paris/FR

Erstanmelder: Rainer Martin, Aachen/DE**Bezeichnung:**Verfahren und Vorrichtung zur Unterdrückung von
Rauschen und Echos**IPC:**

H 03 H, G 10 L, H 04 B

Die angehefteten Stücke sind eine richtige und genaue Wiedergabe der ur-
sprünglichen Unterlagen dieser Patentanmeldung.

München, den 25. September 2000
Deutsches Patent- und Markenamt
Der Präsident
Im Auftrag

Dzierzon

Verfahren und Vorrichtung zur Unterdrückung von Rauschen und Echos

Die Erfindung betrifft ein Verfahren zur Rausch- und Echounterdrückung in Signalen, z.B. Bild- oder Sprachsignalen. Die Erfindung kann zum Beispiel in einer Freisprecheinrichtung aufweisenden Fernsprecheinrichtung, einem Bildtelefon oder einem medizinischen bildgebenden Gerät eingesetzt werden. Die Erfindung betrifft desweiteren eine Vorrichtung zur Ausführung dieses Verfahrens.

Bei der Aufnahmen und der Übertragung von Bild- oder Sprachsignalen werden diese Signale oft durch Rauschen und bei der Übertragung über Leitungen oft auch durch (Leitungs-) Echos gestört. Bei Sprachsignalen und der Verwendung einer Freisprecheinrichtung kommen desweiteren auch noch akustische Echos hinzu. Zur Verbesserung der subjektiven Qualität und der Verständlichkeit bei Sprache, bzw. der Schärfe bei Bildern, können diese Störungen mit einer Störgeräuschreduktion und gegebenenfalls einer Echoreduktionsvorrichtung, z.B. einem Echokompensator oder einer Pegelwaage, reduziert werden.

Die Störgeräuschreduktion wird meist mit einem Verfahren der Spektralen Gewichtung in einem transformierten Bereich, z.B. nach einer Fourier Transformation oder einer Diskreten Cosinus-Transformation ausgeführt. Verfahren für die Verbesserung verrauschter Sprachsignale sind z.B. in Y. Ephraim and D. Malah, 'Speech Enhancement Using a Minimum Mean-Square Error Short-Time Spectral Amplitude Estimator', IEEE Trans. Acoustics, Speech and Signal Processing, Vol. 32, pp. 1109-1121, 1984, und in D. Malah, R.V. Cox and A.J. Accardi, 'Tracking Speech-Presence Uncertainty to Improve Speech Enhancement in Non-Stationary Noise Environments', Proc. IEEE Intl. Conf. Acoustics, Speech, Signal Processing (ICASSP), 1999, beschrie-

ben. Für die Reduktion von Rauschen in Bildsignalen sind ähnliche Verfahren bekannt, siehe z.B. J.S. Lim, 'Image Restoration by Short Space Spectral Subtraction', IEEE Trans. Acoustic, Speech, and Signal Proc., Vol. 28, pp. 191-197, 1980 oder T. Aach and D. Kunz, Spectral Estimation Filters for Noise Reduction in X-Ray Fluoroscopy Imaging', Proc. EUSIPCO, pp. 571-574, 1996.

Die Reduktion von akustischen Echos oder Leitungsechos wird meist mit einem Echokompensator, einer Pegelwaage, einem Center Clipper oder beliebigen Kombinationen dieser Verfahren durchgeführt. Eine Übersicht zum Stand der Technik im Hinblick auf Sprachsignale ist z.B. in E. Hänsler, 'The Hands-Free Telephone Problem - An Annotated Bibliography', Signal Processing, Vol. 27, pp. 259-271, 1992, und in E. Hänsler, 'The Hands-Free Telephone Problem - An Annotated Bibliography Update, Annales des Télécommunication, Vol. 49, No. 7-8, pp. 360-367, 1994, gegeben.

Seit einiger Zeit sind auch Verfahren bekannt, die Rauschen und Echos gemeinsam reduzieren. Ein solches Verfahren ist z.B. in S. Gustafsson et.al.: Combined Acoustic Echo Control and Noise Reduction for Hands-free Telephony, Signal Processing, vol. 64, pp. 21-32, 1998, beschrieben. Diese Verfahren bestehen meist aus der Kombination eines Echokompensators mit einem die Restechos und das Rauschen vermindern dem Filter.

Der Nachteil der bisher bekannten Verfahren besteht darin, dass sie entweder das Nutzsignal (z.B. Sprache) zu stark verzerren, oder aber das verbleibende Hintergrundgeräusch bei Sprache unnatürlich klingt oder stark fluktuiert. In der Bildverarbeitung liegen die Probleme in Blockeffekten, Kontrastverlusten oder aber unnatürlich wirkende Hintergrundmustern ('gratings').

Die Aufgabe der Erfindung besteht deshalb darin, ein Verfahren und eine Vorrichtung anzugeben, die eine wirkungsvolle Rausch- und/oder Echoreduktion ermöglicht, und dabei nur zu minimalen Verzerrungen im Nutzsignal und unwesentlichen Fluktuationen im Reststörsignal führt.

Diese Aufgabe wird durch ein Verfahren gelöst, dass die Merkmale des Anspruchs 1 aufweist.

Das Verfahren beruht darauf, dass das gestörte Signal Nutzsignal, Rauschen und die evtl. auftretenden Echosignale nicht zu allen Zeitpunkten oder in allen Bildpunkten oder Bildregionen enthält und daher der Einsatz unter-

schiedlicher Signalschätzer oder Signalfilter zu unterschiedlichen Zeitpunkten oder bei unterschiedlichen Bildpunkten von Vorteil ist. Beispielsweise ist das Nutzsignal nur dann in dem gestörten Signal enthalten, wenn tatsächlich gerade Bildinformation oder Sprache übertragen wird. In gleicher Weise hängt das Auftreten von Echos von der Aktivität der Signalquellen ab.

Erfindungsgemäss werden nun für die verschiedenen Fälle, z.B. 'Bild- oder Sprachinformation im Signal vorhanden', 'Rauschen im Signal vorhanden', 'Echos im Signal vorhanden', und beliebigen Kombinationen derselben und je nach Pegel der Signal relativ zueinander, unterschiedliche Signalschätzer oder Signalfilter eingesetzt, wobei auch beliebige lineare oder nicht-lineare Kombinationen dieser Signalschätzer oder Signalfilter möglich sind. Auf diese Weise wird das Nutzsignal aus dem gestörten Signal mit einem der jeweiligen Situation angepassten Schätzer oder Filter extrahiert. Das Umschalten dieser Schätzer oder Filter oder ihre relative Gewichtung bei einer Kombination wird mit Hilfe von Variablen bestimmt, die aus dem gestörten Signal oder aus anderen Signalquellen bestimmt werden.

Fig. 1 zeigt die prinzipielle Anordnung. $z(k)$ bezeichnet ein Sprach- oder Bildsignal und $\hat{s}(k)$ das von Rauschstörungen oder Echos befreite Signal. $x(k)$ bezeichnet ein möglicherweise vorhandenes externes Steuersignal. Das gestörte Signal $z(k)$ wird nun N Signalschätzern oder Signalfiltern S_1, S_2, \dots, S_N zugeführt. Eine Kombinationseinheit K , die das Eingangssignal $z(k)$ und das evtl. vorhandenen externen Steuersignal $x(k)$ verarbeitet, steuert die Auswahl und/oder die Kombination der von den Signalschätzern oder Signalfiltern S_1, S_2, \dots, S_N gelieferten Signale und bildet das von Rauschen und Echos befreite Signal $\hat{s}(k)$.

Zur Bestimmung der Schätzer und der relativen Gewichtung dieser Schätzer zueinander sind besonders statistische Verfahren geeignet. Insbesondere kann die Auswahl eines oder mehrerer zum einem bestimmten Zeitpunkt oder in einem bestimmten Bildpunkt geeigneter Signalschätzer oder Signalfilter aus den geschätzten Wahrscheinlichkeiten für das Vorhandensein der Nutzsignal, Rausch- oder Echosignalanteile im gestörten Signal gesteuert werden. Nachfolgend wird ein Ausführungsbeispiel eines solchen Schätzverfahrens für den Einsatz in einer Freisprecheinrichtung angegeben.

Fig. 2 zeigt die Kombination aus einem Echokompensator und einem adaptiven Nachfilter. Das Nachfilter ist im Frequenzbereich, unter Verwendung

einer Diskreten Fouriertransformation (DFT), einer spektralen Gewichtung, und einer 'overlap/add'-Signalsynthese implementiert. Das Nachfilter setzt sich, wie im folgenden gezeigt wird, aus zwei Teilfiltern und einer adaptiven Kombinationseinheit zusammen.

Wir betrachten die abgetasteten und bandbegrenzten Signale $x(i)$, $y(i)$, $z(i)$ und $\hat{s}(i)$ wobei i den diskreten Zeitindex bezeichnet. $x(i)$ ist das Signal des fernen Sprechers und $y(i)$ ist das Mikrophonsignal welches sich aus einem Sprachsignal $s(i)$, einem Rauschen $n(i)$, und einem Echosignal $e(i)$, $y(i) = s(i) + n(i) + e(i)$, zusammensetzt. Das echokompensierte Signal $z(i)$ ist das Mikrophonsignal minus dem vom Kompensator geschätzten Echo $\hat{e}(i)$, $z(i) = y(i) - \hat{e}(i) = s(i) + n(i) + \tilde{e}(i)$. $\tilde{e}(i)$ bezeichnet das Restecho nach der Kompensation. Wir nehmen ferner an, dass die Signale $s(i)$, $x(i)$, und $n(i)$ statistisch unabhängig sind. Das durch Rauschen und Restechos gestörte kompensierte Signal $z(i)$ wird unter Verwendung einer Fensterfunktion $h(i)$ in den Frequenzbereich transformiert, indem ein Rahmen von L aufeinanderfolgende Abtastwerte von $z(i)$ zusammengefasst werden, dieser Rahmen mit der Fensterfunktion gewichtet wird und eine DFT der Länge L berechnet wird. Vor der nächsten DFT Berechnung wird das Fenster um R Abtastwerte auf dem Eingangssignal verschoben. Die DFT Analyse mit gleitendem Fenster resultiert in einer Menge von Frequenzbereichssignalen, die auch mit

$$Z(\lambda, k) = \sum_{\mu=0}^{L-1} z(\lambda R - \mu) h(\mu) e^{-j2\pi k\mu/L} \quad (1)$$

angegeben werden können, wobei λ ein unterabgetasteter Zeitbereichsindex, $\lambda \in \mathbb{Z}$, und k ein Frequenzindex, $k \in \{0, 1, \dots, L-1\}$, angibt, und letzterer zur normierten Mittenfrequenz der DFT-Bänder mit Ω_k by $\Omega_k = k2\pi/L$ in Beziehung steht. Typischerweise wird eine Abtastrate von $f_A = 8000$ Hz und eine DFT-Länge von $L = 2R = 256$ verwendet.

Ebenso werden die Fourierkoeffizienten aller anderen Signale des k ten Frequenzindex mit

- $S_k = A_k \exp(j\alpha_k)$ (ungestörte nahe Sprache),
- $Y_k = R_k \exp(j\vartheta_k)$ (gestörte nahe Sprache),
- $X_k = B_k \exp(j\beta_k)$ (ferne Sprache),
- $Z_k = D_k \exp(j\zeta_k)$ (kompensiertes Signal),

- $\hat{S}_k = \hat{A}_k \exp(j\hat{\alpha}_k)$ (entstörtes Signal).

angegeben, wobei der Klarheit halber der Zeitindex λ unterdrückt wurde.

Die Verteilungsdichtefunktionen der nahen und der fernen Sprache können mit $p_s(A_k, \alpha_k)$, bzw. $p_x(B_k, \beta_k)$, angegeben werden:

$$p_s(A_k, \alpha_k) = P(H_{sk}^1)p_s(A_k, \alpha_k|H_s^1) + P(H_{sk}^0)\delta(A_k, \alpha_k) \quad (2)$$

$$p_x(B_k, \beta_k) = P(H_{xk}^1)p_x(B_k, \beta_k|H_x^1) + P(H_{xk}^0)\delta(B_k, \beta_k) \quad (3)$$

wobei $P(H_{sk}^1)$ und $P(H_{xk}^1)$ für die Wahrscheinlichkeiten, dass nahe bzw. ferne Sprache vorliegt, stehen und $P(H_{sk}^0) = 1 - P(H_{sk}^1)$ und $P(H_{xk}^0) = 1 - P(H_{xk}^1)$ gilt. $\delta(\cdot)$ bezeichnet die Dirac-Funktion.

Es wird nun die Kostenfunktion

$$\begin{aligned} \mathcal{C} = & \int_{\Omega_Z} \{ P(H_s^0)P(H_x^0)\hat{A}^2 p(Z|H_s^0, H_x^0) \\ & + P(H_s^1)P(H_x^1) \iiint (\hat{A} - A)^2 p(Z|A, \alpha, B, \beta) \\ & \cdot p_s(A, \alpha|H_s^1)p_x(B, \beta|H_x^1) dA d\alpha dB d\beta \\ & + P(H_s^0)P(H_x^1)\hat{A}^2 p(Z|H_s^0, H_x^1) \\ & + P(H_s^1)P(H_x^0) \iint (\hat{A} - A)^2 \\ & \cdot p(Z|A, \alpha, H_x^0)p_s(A, \alpha|H_s^1) dA d\alpha \} dZ \end{aligned} \quad (4)$$

minimiert, wobei \hat{A} die spektralen Amplituden des geschätzten nahen Signals bezeichnen, die wiederum eine Funktion von Z sind. Die Lösung ist durch

$$\begin{aligned} \hat{A} = & \frac{P(H_s^1)P(H_x^1)p(Z|H_s^1, H_x^1)}{P_\Sigma} S_1 \\ & + \frac{P(H_s^1)P(H_x^0)p(Z|H_s^1, H_x^0)}{P_\Sigma} S_2 \end{aligned} \quad (5)$$

gegeben, mit

$$S_1 = E\{A|Z, H_s^1, H_x^1\} \quad (6)$$

$$S_2 = E\{A|Z, H_s^1, H_x^0\} \quad (7)$$

gekennzeichnet, dass es sich bei den Echos um akustische Echos handelt;

11. Ein Verfahren nach einem der vorangegangenen Ansprüche, dadurch gekennzeichnet, dass es sich bei den Echos um Leitungsechos handelt;

12. Ein Verfahren nach einem der vorangegangenen Ansprüche, dadurch gekennzeichnet, dass es sich bei den Echos um Echos in einer Mess- oder Signalaufnahmeverrichtung handelt;

13. Ein Verfahren nach einem der vorangegangenen Ansprüche, dadurch gekennzeichnet, dass die Echos überwiegend von einem Echokompensator reduziert werden und die Restechos und das Rauschen von nachgeschalteten Signalschätzern oder Signalfiltern reduziert werden.

14. Ein Verfahren nach einem der vorangegangenen Ansprüche, dadurch gekennzeichnet, dass die Signalschätzer oder Signalfilter in einem transformierten Bereich, z.B. im Bereich einer Fourier- oder Diskreten Cosinustransformation, berechnet werden;

15. Ein Verfahren nach einem der vorangegangenen Ansprüche, dadurch gekennzeichnet, dass die Signaltransformation eine diskrete Fouriertransformation ist;

16. Ein Verfahren nach einem der vorangegangenen Ansprüche, dadurch gekennzeichnet, dass die Signaltransformation in Frequenzrichtung nach einer Bark oder Mel Skala ausgeführt wird;

17. Ein Verfahren nach einem der vorangegangenen Ansprüche, dadurch gekennzeichnet, dass die Größen zur Steuerung der Signalkombination aus den Wahrscheinlichkeiten für das Vorhandensein des Signals, des Rauschens und des Echos berechnet werden;

18. Ein Verfahren nach einem der vorangegangenen Ansprüche, dadurch gekennzeichnet, dass die Signalschätzer oder Signalfilter nach dem Prinzip der kleinsten Fehlerquadrate berechnet werden;

19. Ein Verfahren nach einem der vorangegangenen Ansprüche, dadurch gekennzeichnet, dass die Signalschätzer oder Signalfilter nur die Amplituden der Signale in einem transformierten Bereich berücksichtigen;

20. Ein Verfahren nach einem der vorangegangenen Ansprüche, dadurch gekennzeichnet, dass die Signalschätzer oder Signalfilter gemeinsam eine

Kostenfunktion minimieren;

21. Ein Verfahren nach einem der vorangegangenen Ansprüche, **dadurch gekennzeichnet, dass** die Ausgangssignale der Signalschätzer oder Signalfilter gewichtet und gemittelt werden und die Gewichte aus den Wahrscheinlichkeiten für das Vorhandensein des Nutzsignals oder der Störsignale berechnet werden;

22. Ein Verfahren nach einem der vorangegangenen Ansprüche, **dadurch gekennzeichnet, dass** das von Rauschen und Echos befreite Signal nach den Gleichungen 1-12 berechnet wird;

23. Ein Verfahren nach einem der vorangegangenen Ansprüche, **dadurch gekennzeichnet, dass** das gleichzeitig akustische Echos und Rauschen unterdrückt wird und die Kombination der Signalschätzer oder Signalfilter von der Aktivität des nahen und des fernen Sprechers einer Freisprecheinrichtung abhängt.

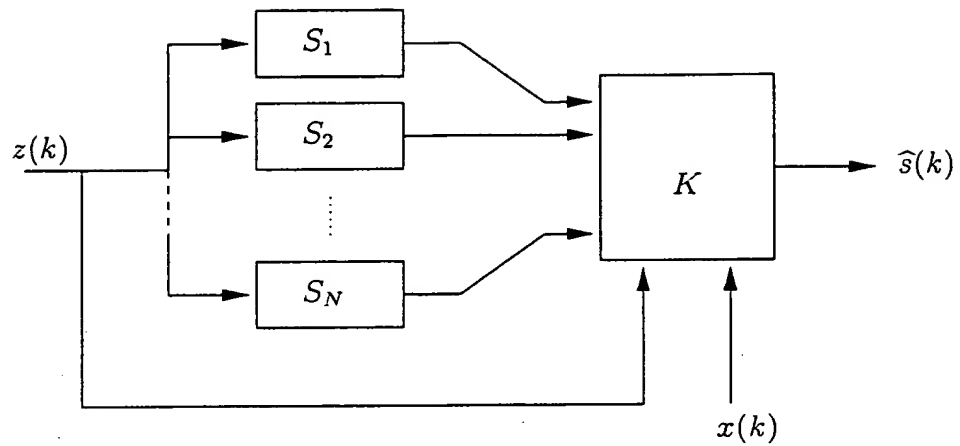


Fig. 1: Vorrichtung zur Reduktion von Rausch- und Echostörungen mit Hilfe von N Signalschätzern oder Signalfiltern und einer Kombinationseinheit K .

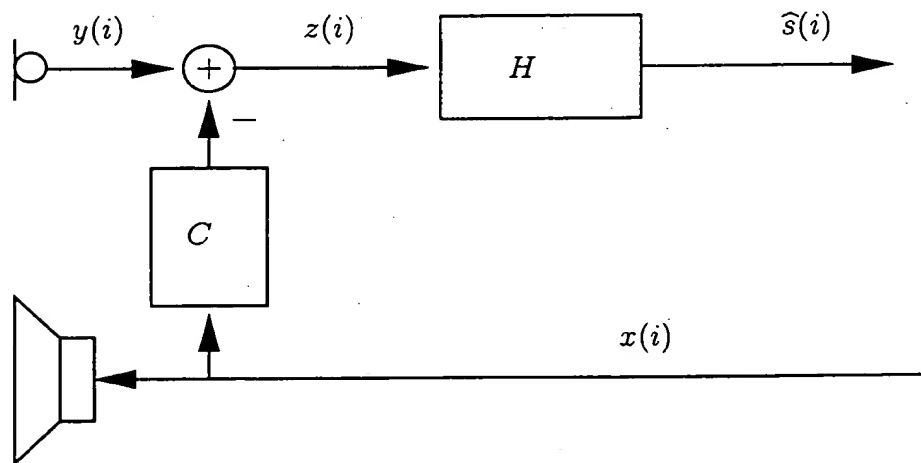


Fig. 2: Vorrichtung zur gemeinsamen Reduktion akustischer Echos und Rauschen.

Zusammenfassung

Die Erfindung beschreibt ein Verfahren und eine Vorrichtung zur Unterdrückung oder Reduktion von Rauschen und/oder Echos in Sprach- oder Bildsignalen, das aus mehreren Signalschätzern oder Signalfiltern S_1, S_2, \dots, S_N , und einer Kombinationseinheit K besteht. Die Kombinationseinheit erlaubt die Auswahl eines der geschätzten Signale oder die Kombination der Ausgangssignale der Signalschätzer oder Signalfilter zu einem einzigen verbesserten Signal $\hat{s}(k)$. Gegenüber dem Stand der Technik weist die Erfindung geringer Verzerrungen des Nutzsignals und natürlichere, weniger stark fluktuierende Reststörungen auf.

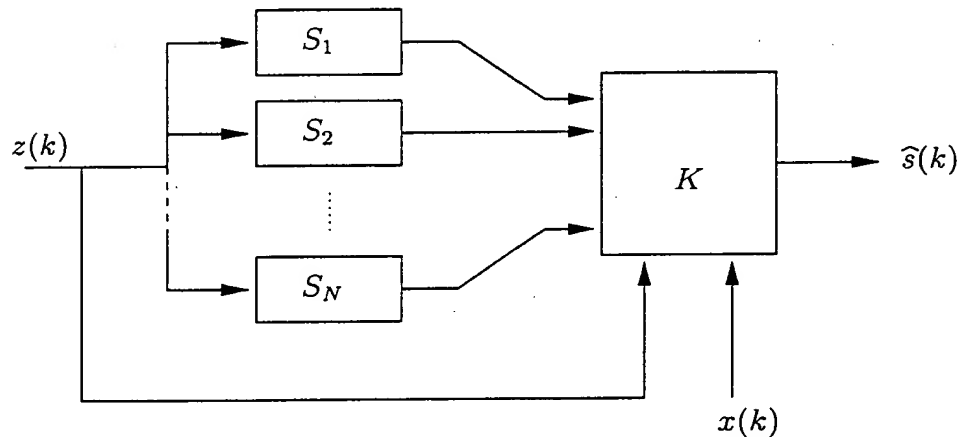


Fig. 1: Vorrichtung zur Reduktion von Rausch- und Echostörungen mit Hilfe von N Signalschätzern oder Signalfiltern und einer Kombinationseinheit K .